HF jet tagging with full G4 tracking simulation

Dennis V. Perepelitsa
Brookhaven National Laboratory

3 May 2016
sPHENIX Simulations Meeting

Reminder of MIE proposal update study

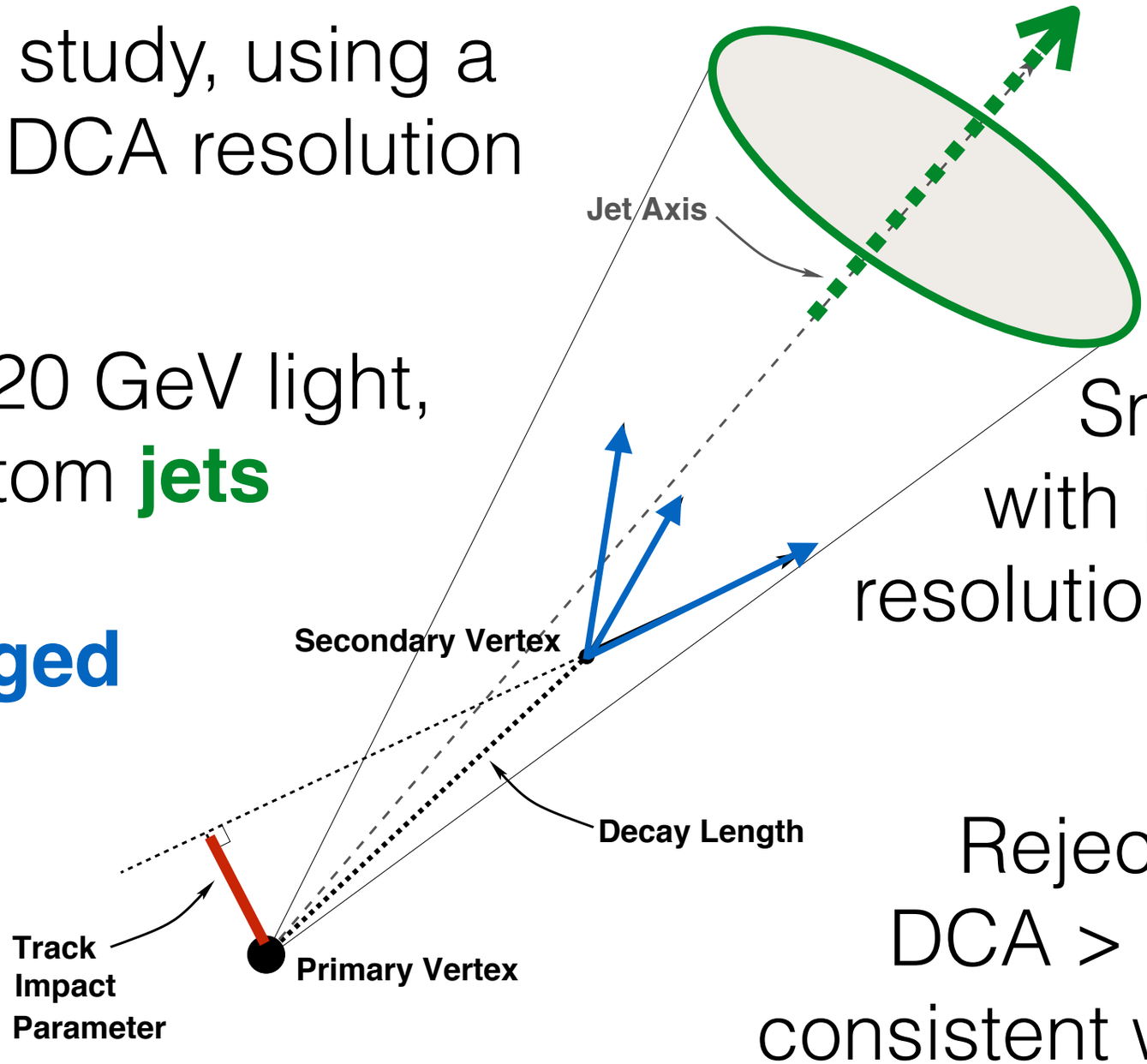
- Explored *b*-jet tagging via "TrackCounting" algorithm, which requires the presence of multiple tracks in the jet with a large impact parameter to PV (or 2-D DCA)
- Parameterized "Fast simulation" study:
  - ➡ used true hadron DCA, smeared by parameterization of DCA resolution from G4 sims
  - ➡ used ideal seven-layer silicon tracker
  - ➡ studied primarily PYTHIA events, with some parameterized UE embedding studies
- Next several slides are from my talk at the DOE Science Review on 30 April 2015

TrackCounting algorithm

Fast simulation study, using a parameterized DCA resolution

Generate $p_T > 20$ GeV light, charm and bottom **jets**

Consider **charged hadrons** with $p_T > 0.5$ GeV in the **jet** cone



Calculate true signed **DCA**

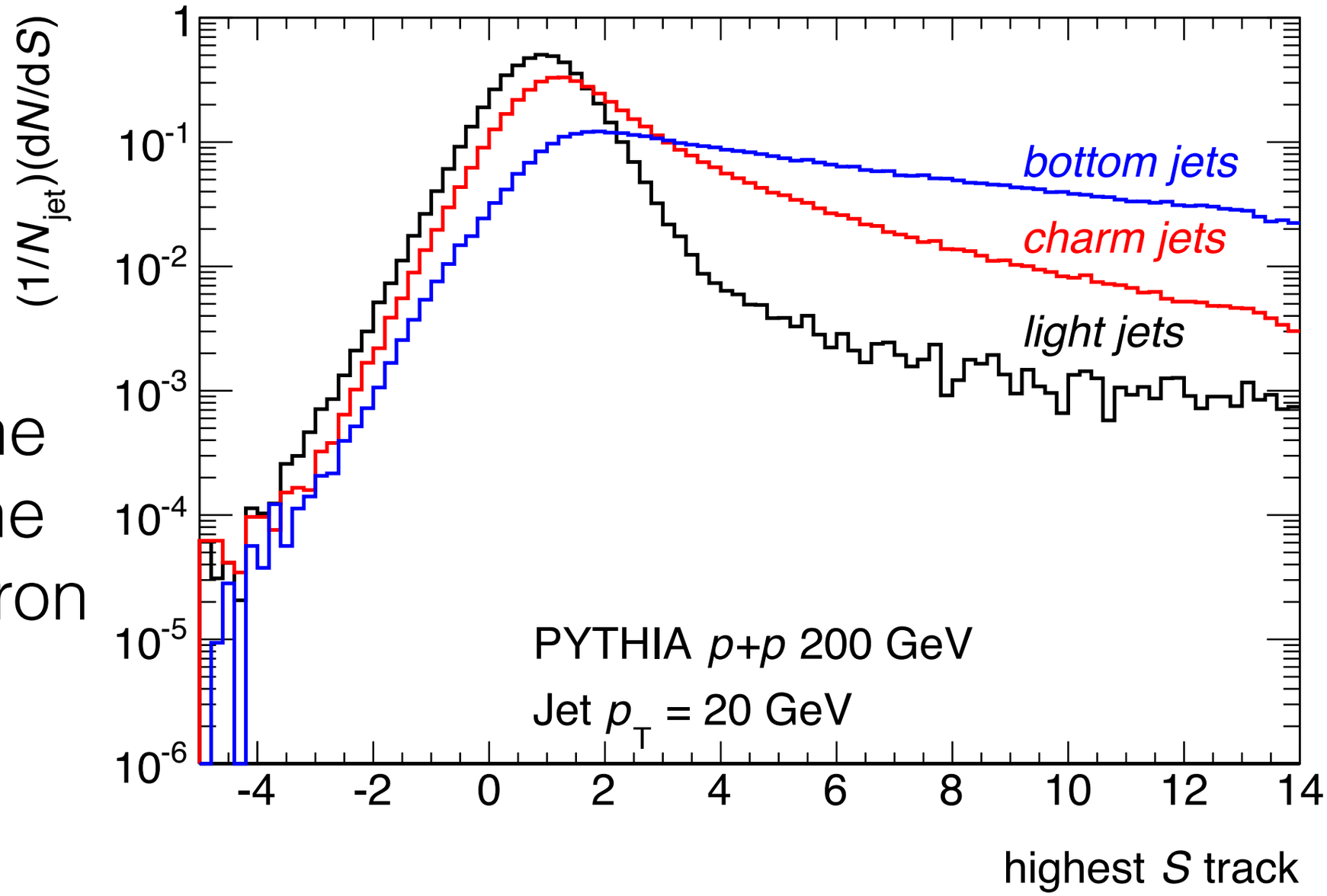
Smear true **DCA** with parameterized resolution derived from GEANT4

Reject hadrons with DCA > 1 mm or those consistent with $V^0 \rightarrow h^{\pm}h^{\pm}$

Sort remaining hadrons by their signed *DCA significance* S_{DCA} = **DCA** / σ_{DCA}

Reconstructed DCA of hadrons in jets

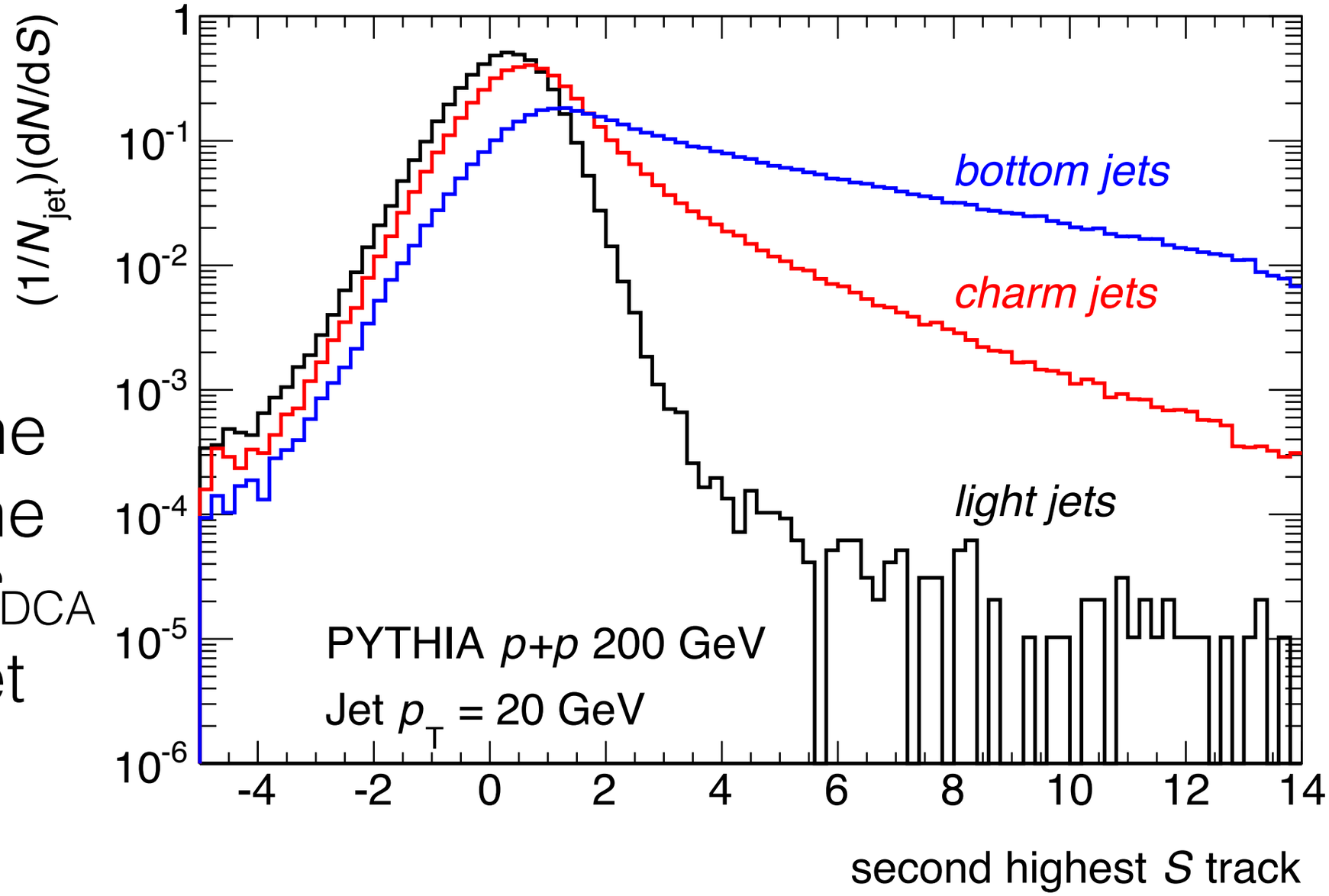
Distribution of the S_{DCA} value for the *highest*-S_{DCA} hadron in the jet



- High-S_{DCA} in **charm**/**bottom** jets from displaced vertices
- High-S_{DCA} tail in **light** jets from tails in DCA resolution and Σ/Ξ decays (cannot be removed with two-track mass analysis)

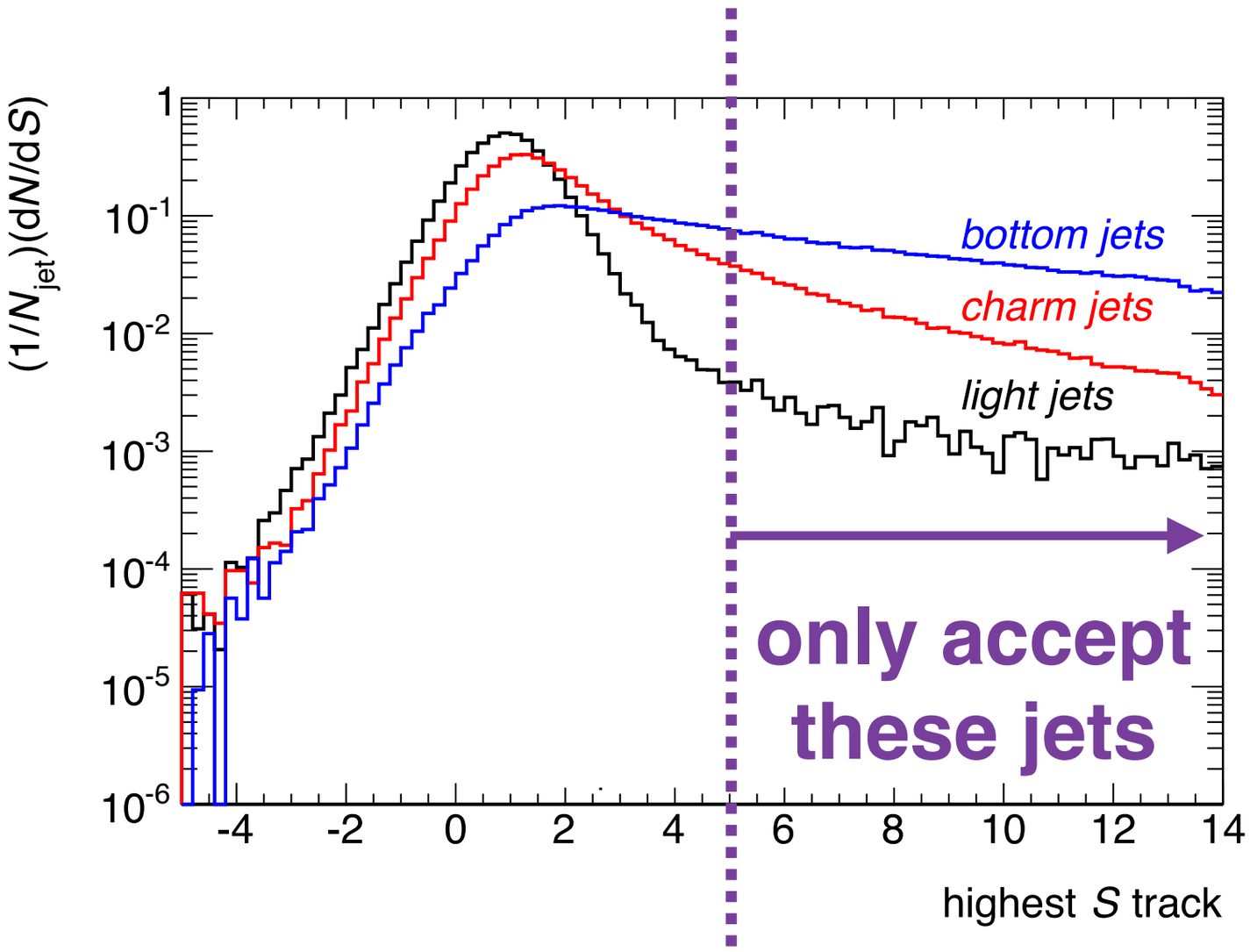
Reconstructed DCA of hadrons in jets

Distribution of the S_{DCA} value for the *second highest*-S_{DCA} hadron in the jet



- Asking for a *second* track with high S_{DCA} cuts down on the **light** jet background
- **Charm** and **bottom** jets retain large-S_{DCA} tails

b-jet performance in *p*+*p*

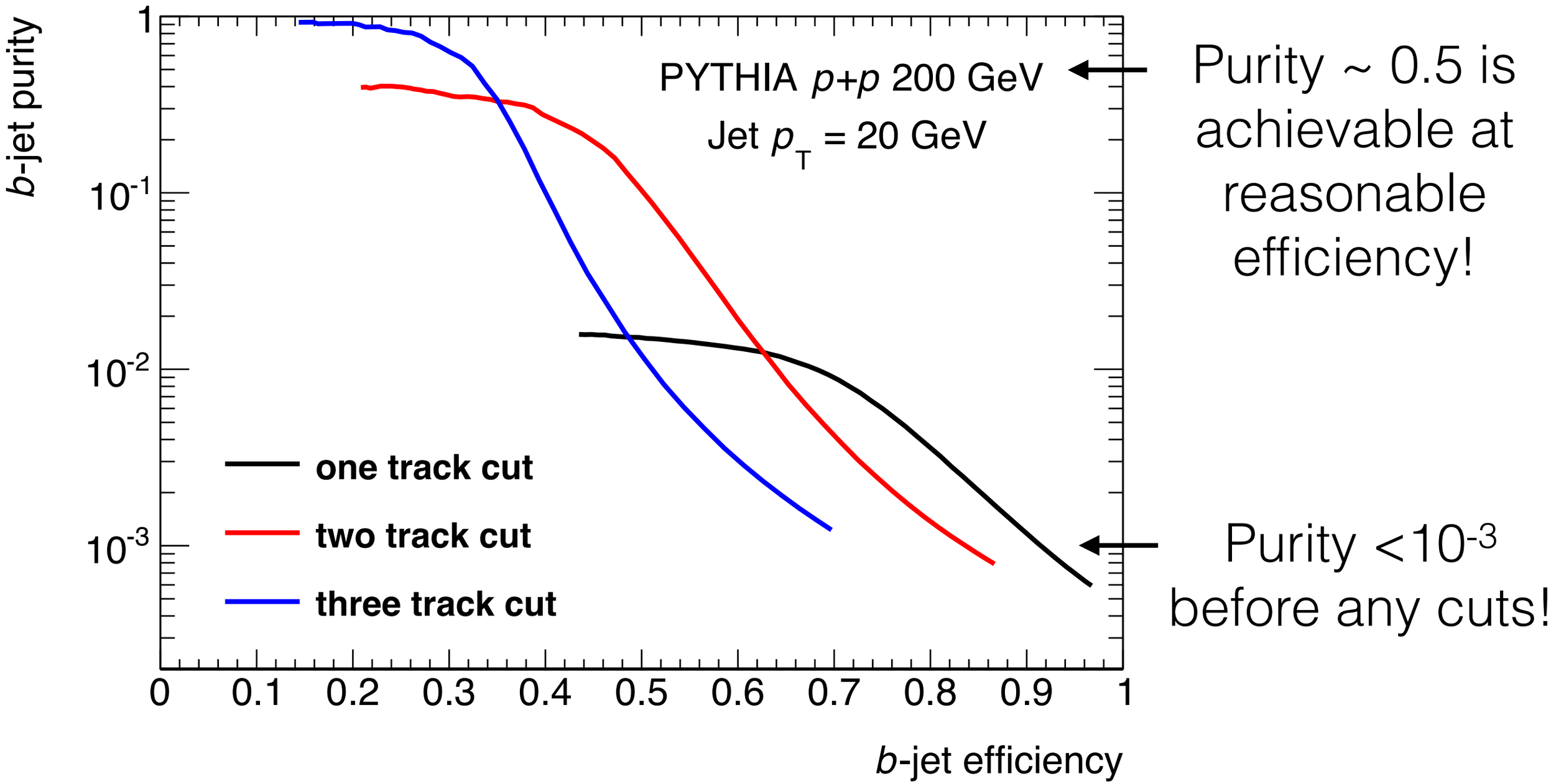


Calculate **efficiency** for light, charm and bottom jets

Calculate **purity** of *b*-jets after cuts as $P = N_b / (N_b + N_c + N_l)$

Vary S_{DCA} cut value to trade off ***b*-jet efficiency** vs. ***b*-jet purity**

b-jet performance in *p*+*p*



Purity ~ 0.5 is achievable at reasonable efficiency!

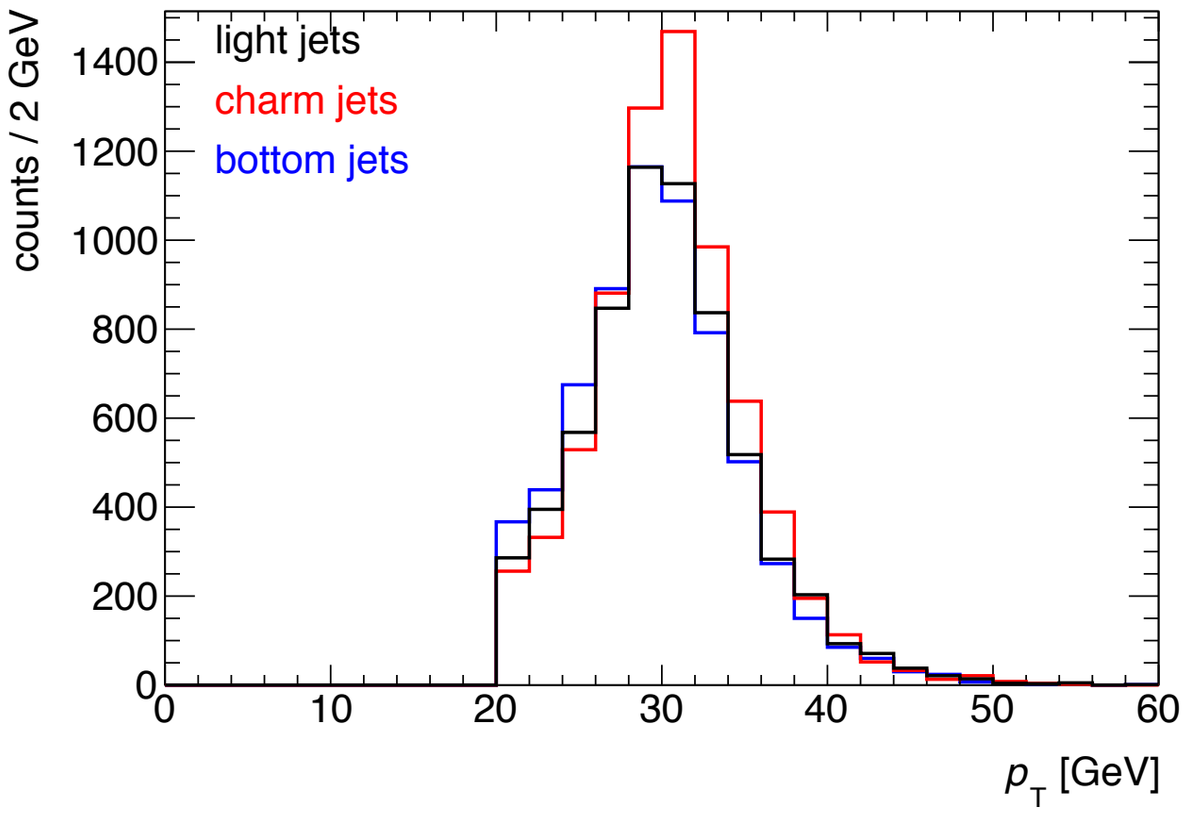
Purity $<10^{-3}$ before any cuts!

P vs. **E** curves for requiring **1**, **2** or **3** tracks with S_{DCA} above some minimum value

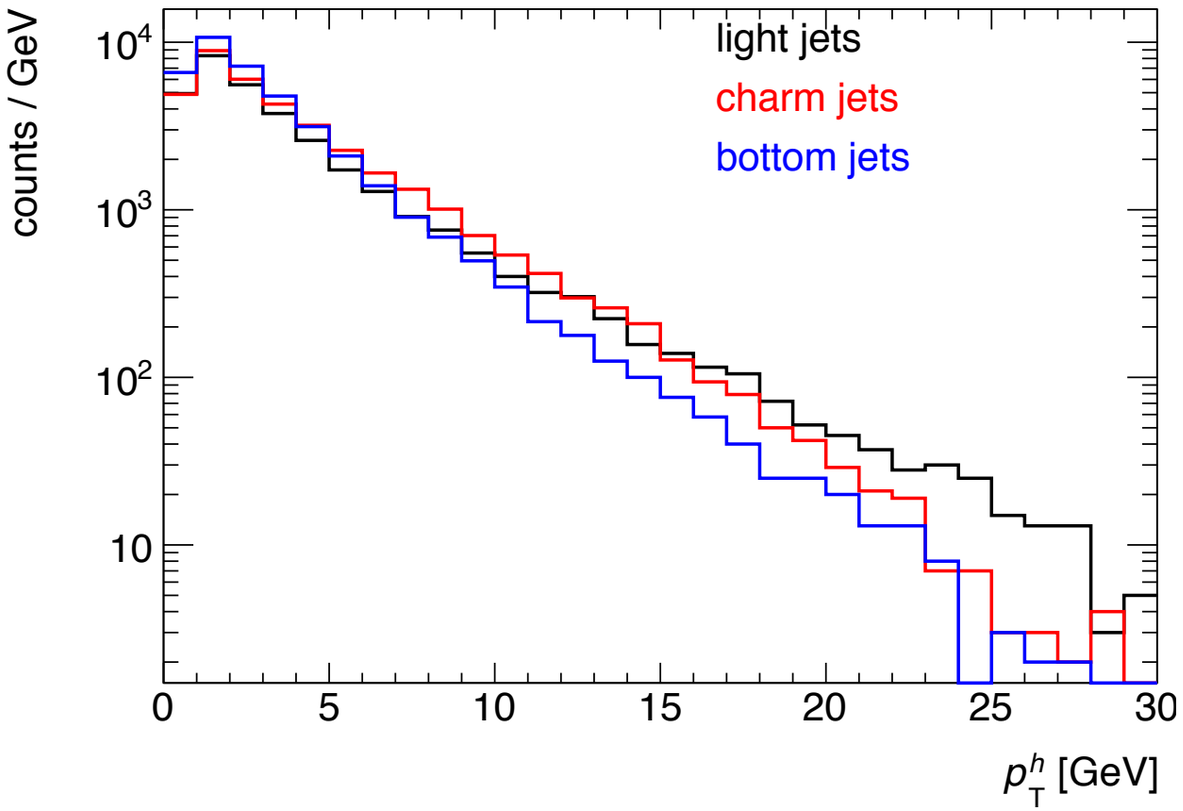
Updated study with G4

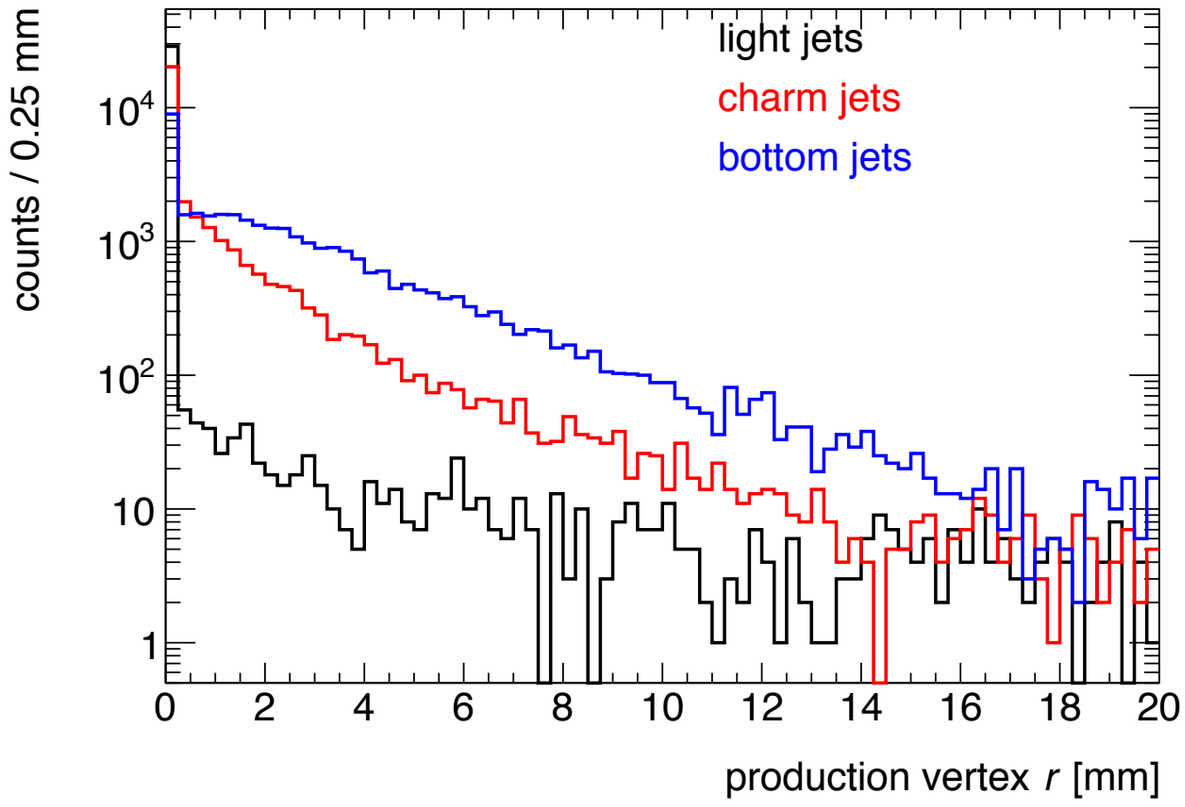
MC sample

- N_{evt} = 10k Pythia8 events with PhaseSpace:pTHatMin = 30.0, with one of the following hard scattering channels on:
  - ➡ HardQCD:all (light), HardQCD:hardccbar (charm), HardQCD:hardbbbar (bottom)
  - ➡ *note*: for more realistic samples, should actually run lots of small-pTHatMin, HardQCD::all events and trigger at truth-level
  - ➡ *note*: "light" sample actually has small admixture of *c*/*b*-jets
- Full G4 tracking simulation, using MIE proposal tracker (ideal seven-layer silicon, G4_Svtx.C)
- Select truth jets with p_T > 20 GeV, $|\eta| < 0.6$
  - ➡ about N_{jet} ~ 13k in each sample
  - ➡ *note*: p_T spectrum unphysical



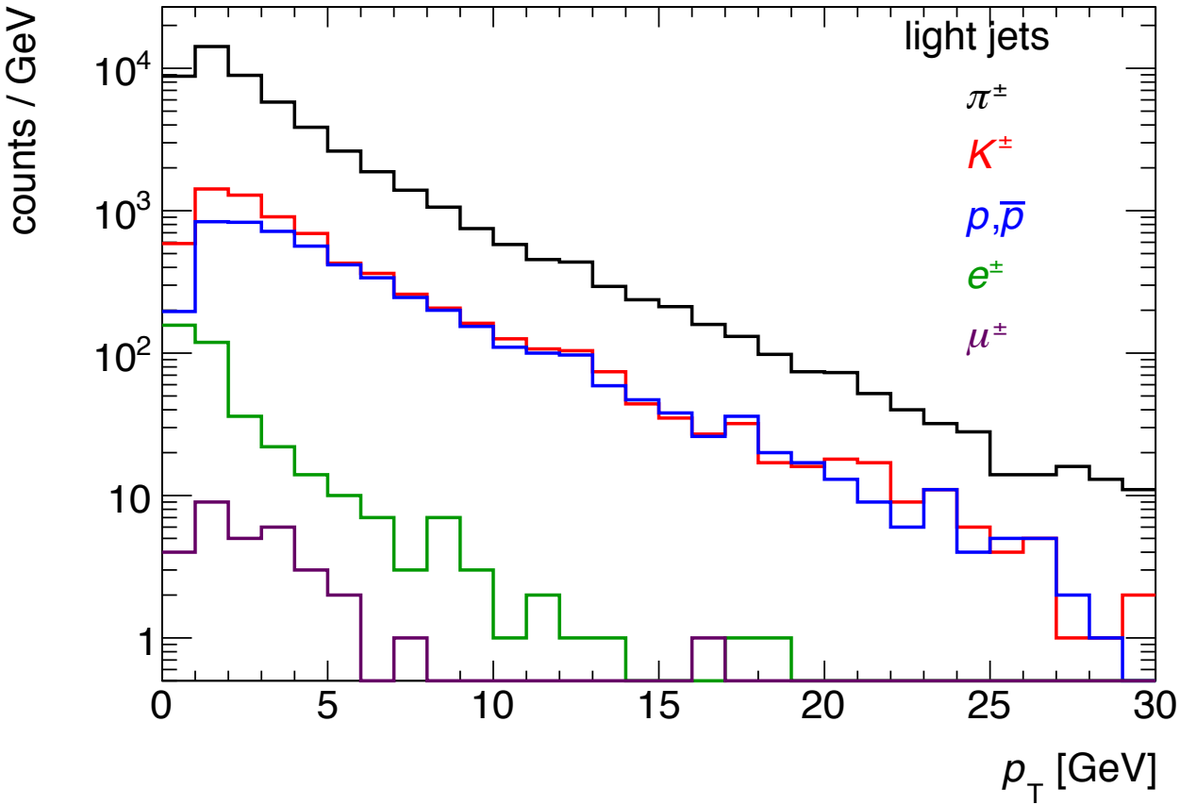
Truth-level p_T and transverse distance from PV

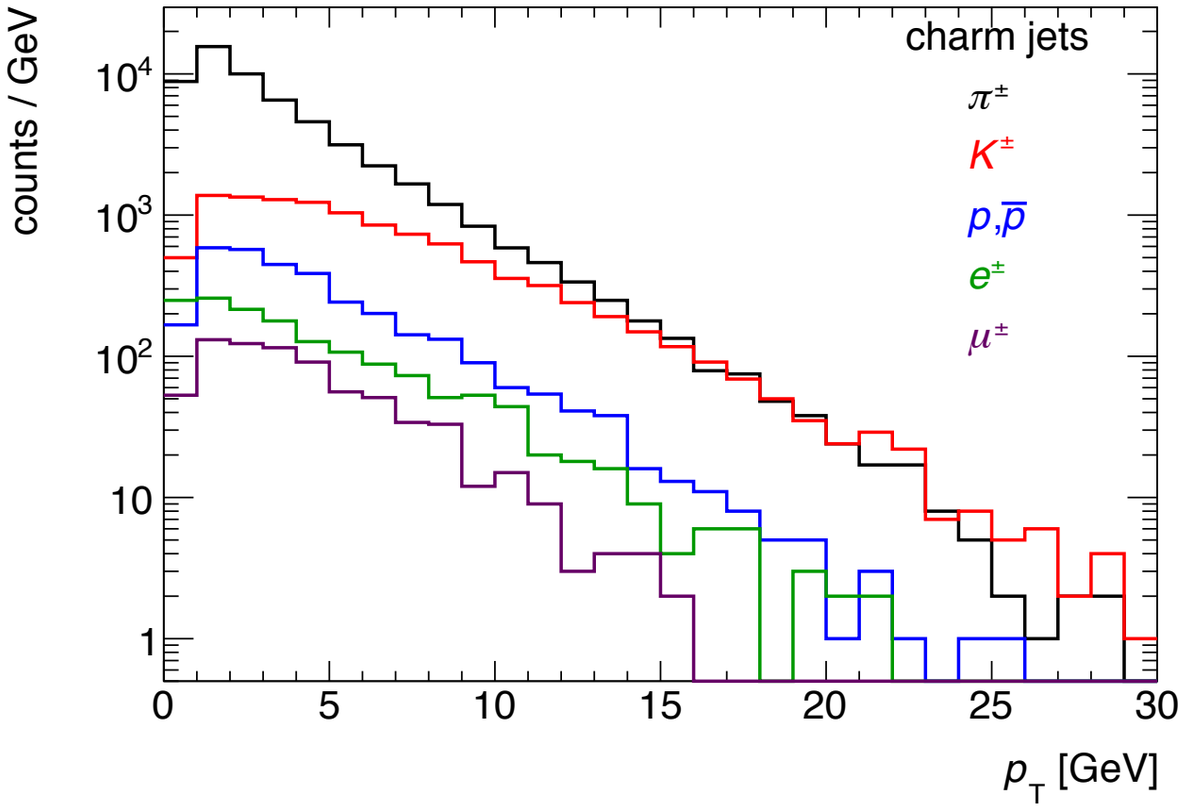


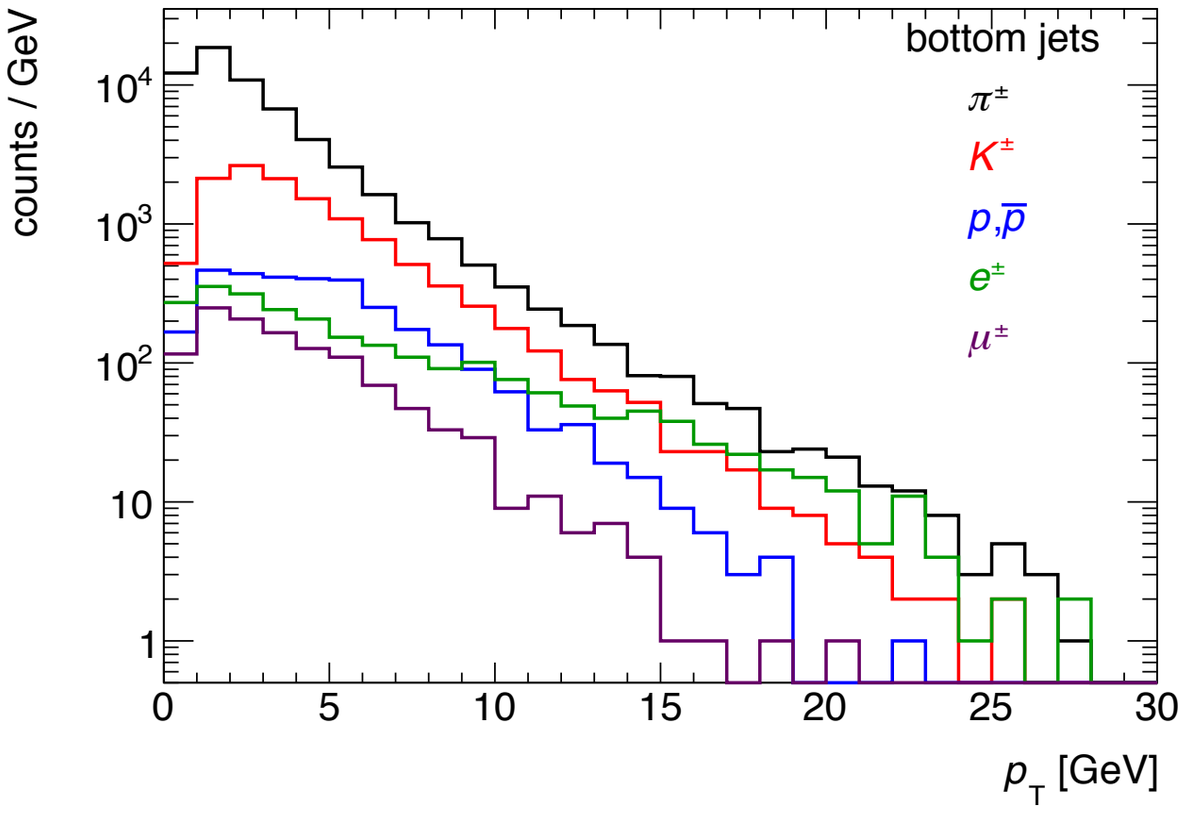


- "Truth-level": analyze final-state charged particles with p_T > 0.5 GeV that are $\Delta R < 0.4$ from the truth jet axis
  - ➡ then, match to HepMC record particle and record $r = \sqrt{(x^2 + y^2)}$ of its production vertex
- r distributions obviously flavor-dependent

Truth-level p_T by species

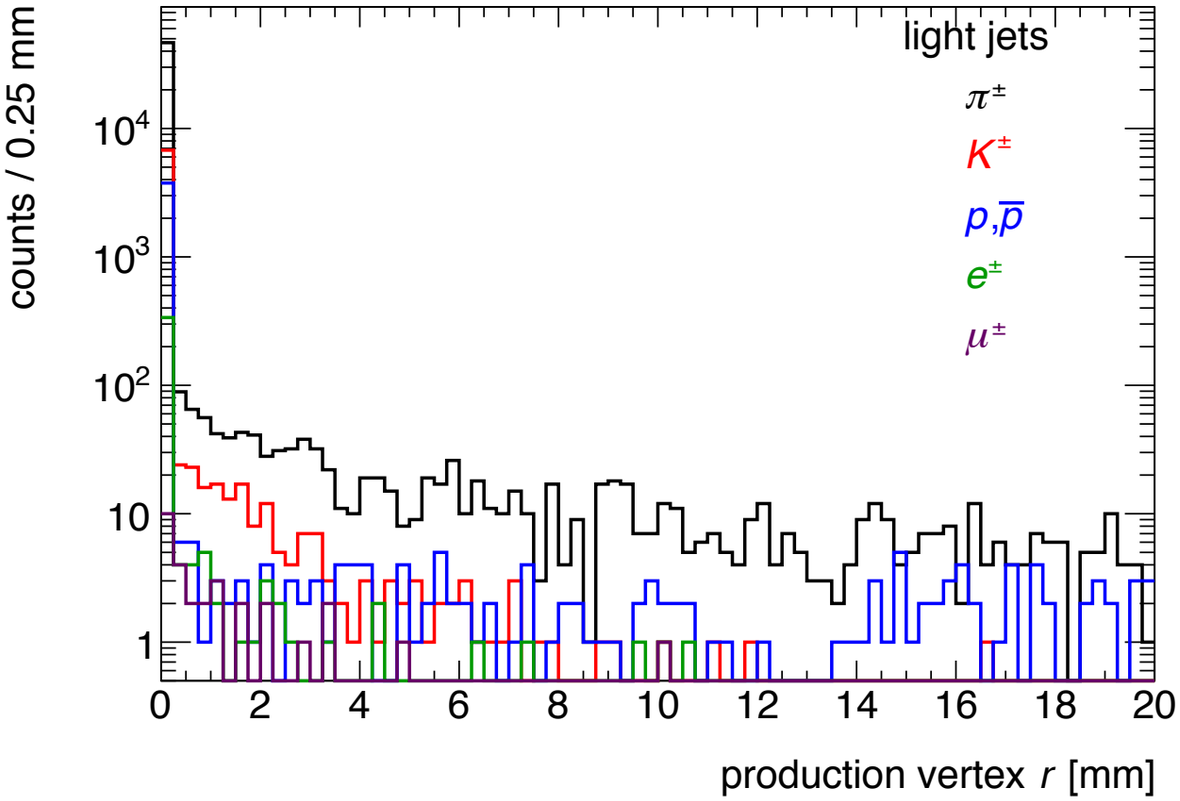


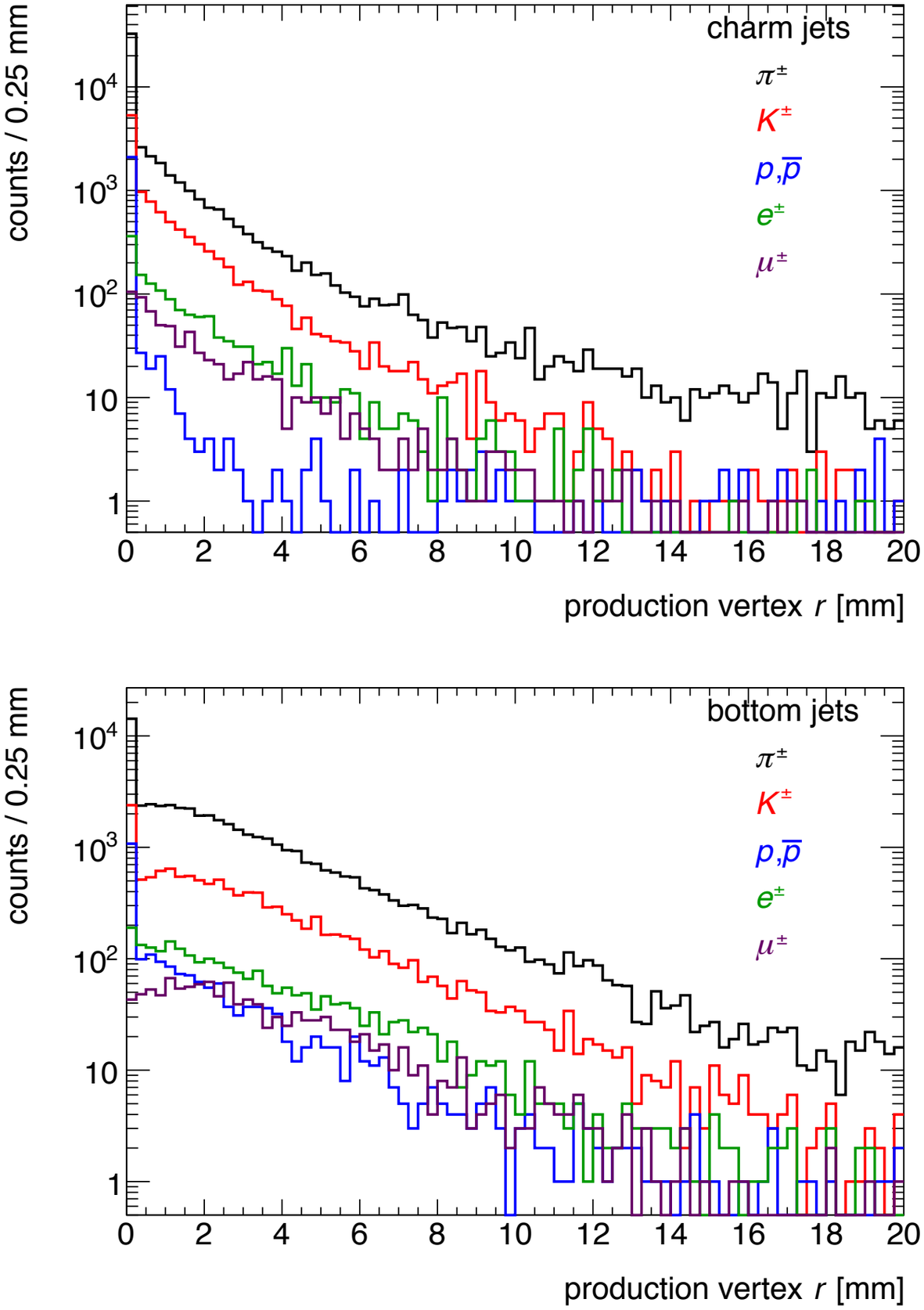




- Each panel is a fixed flavor of jet
  - ➡ showing truth-level p_T spectrum broken out by particle ID
  - ➡ HF jets have more K's, *p*'s, *e*'s and μ's from HF decay chains

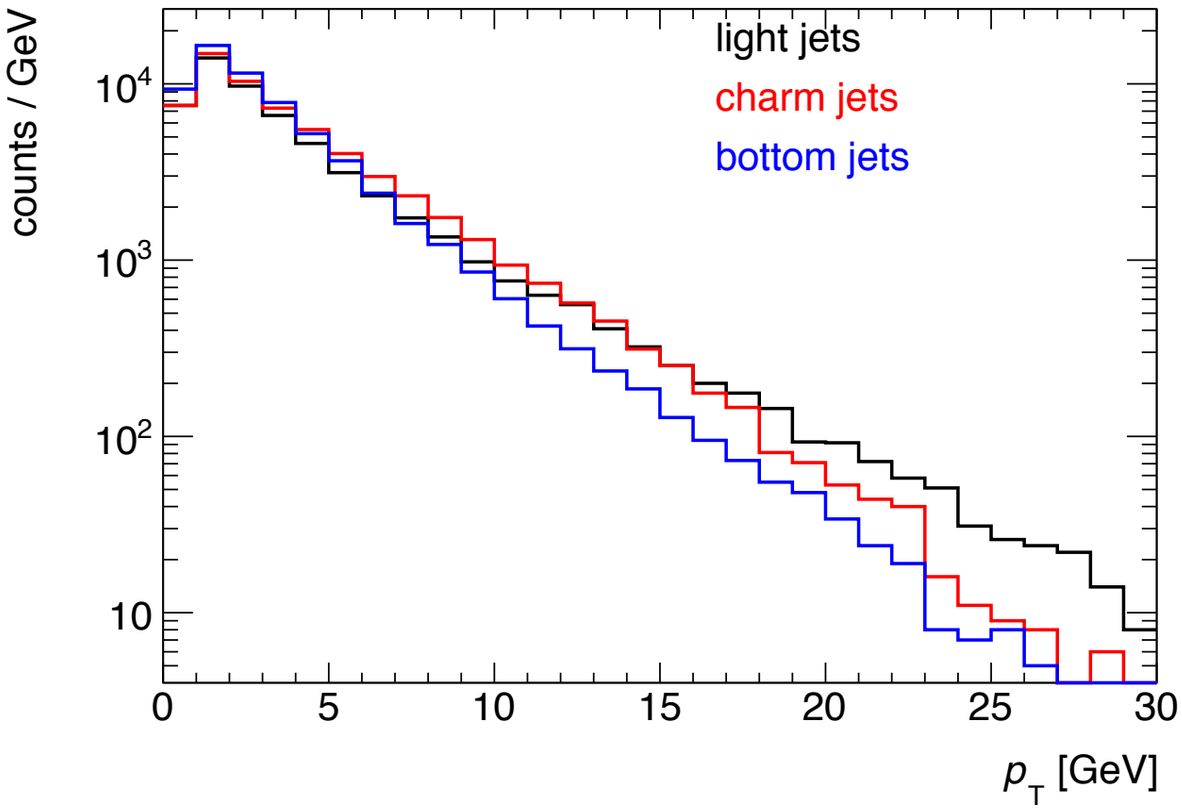
Truth-level production *r* by species

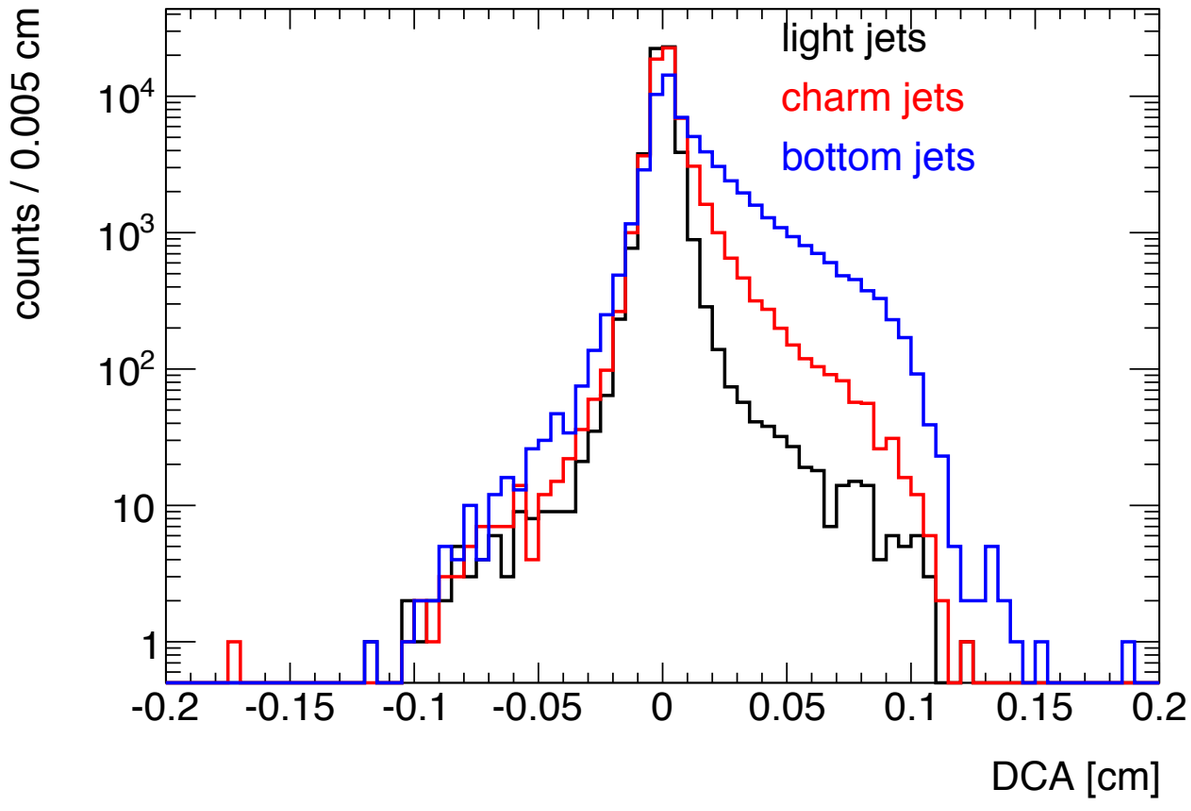




- Each panel is a fixed flavor of jet
  - ➡ showing truth-level *r* distribution broken out by particle ID
  - ➡ particles in HF jets have more displaced production points

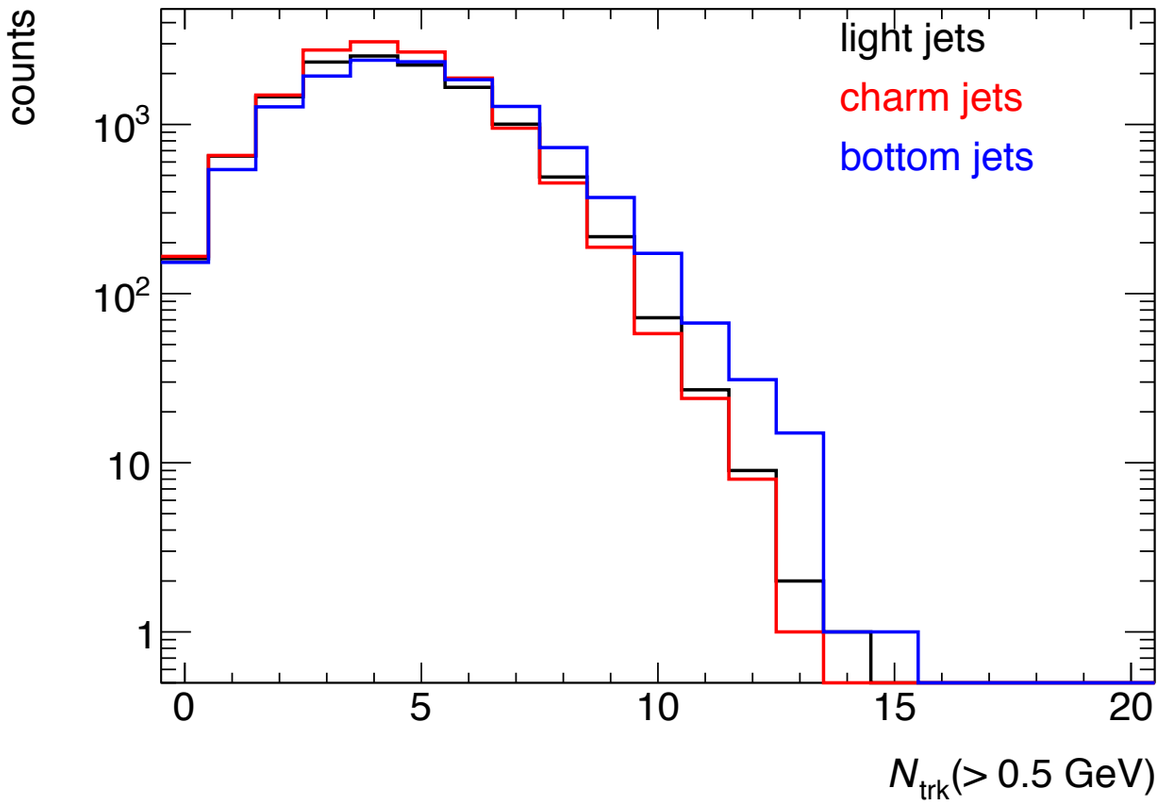
Reco-level p_T and DCA

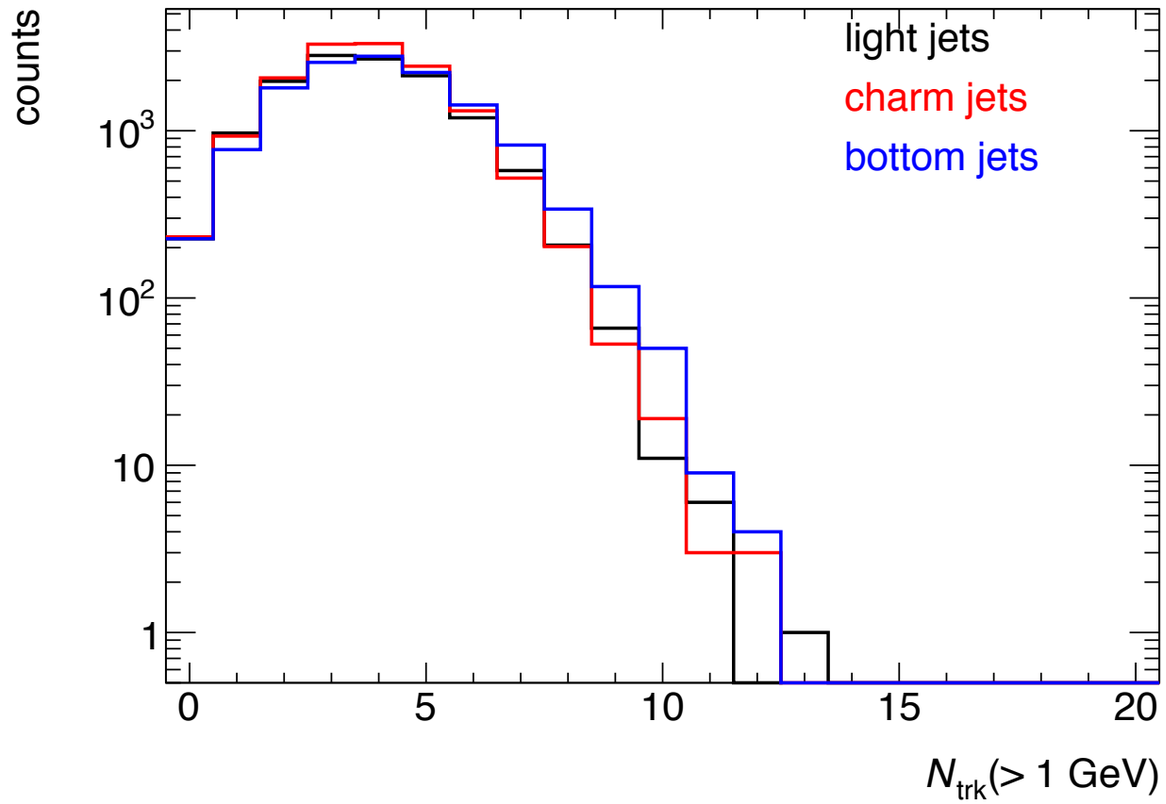




- "Reco-level": tracks with p_T > 0.5 GeV with ΔR < 0.4 to a truth jet
  - ➡ no further selection on $N^{\text{cluster}}_{\text{best}}$, closeness of reco p_T to truth p_T, etc. (i.e. analyzing this as one would analyze data)
- Right: *signed* 2-D DCA distributions for those tracks
  - ➡ ± defined by dot product between DCA vector and jet vector

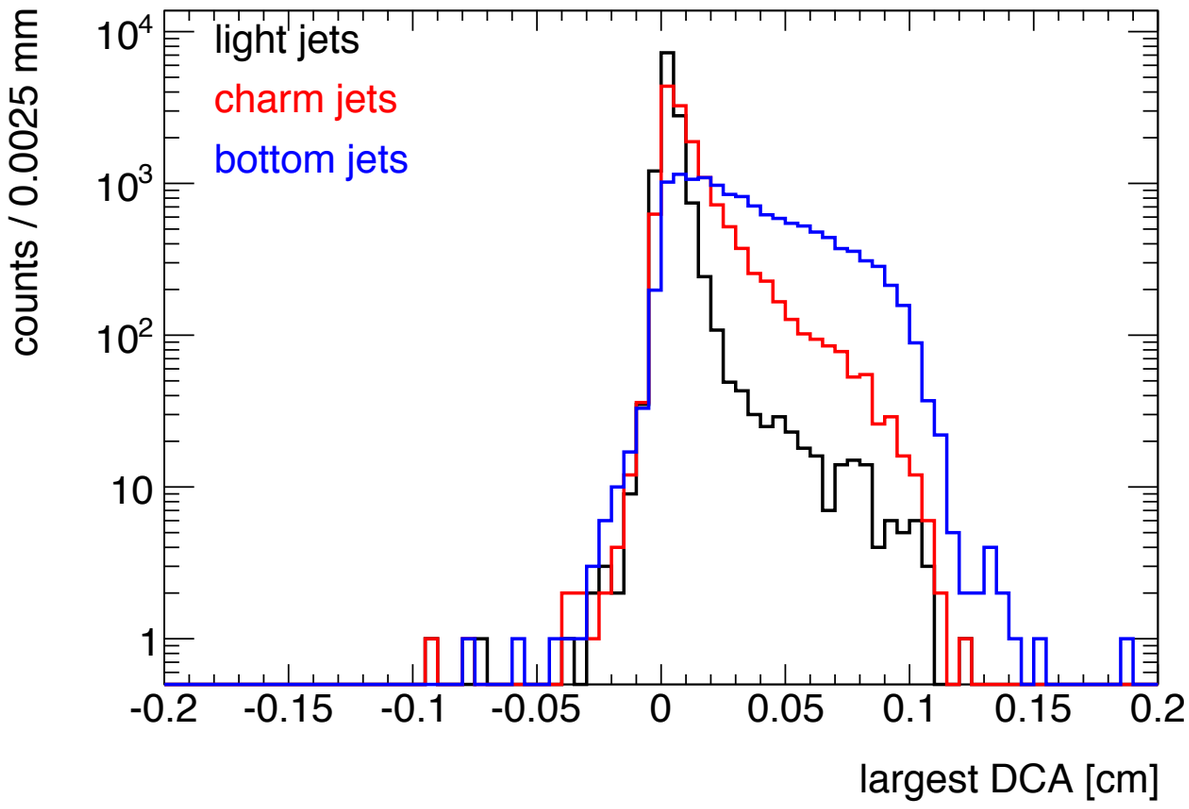
Reco track multiplicities

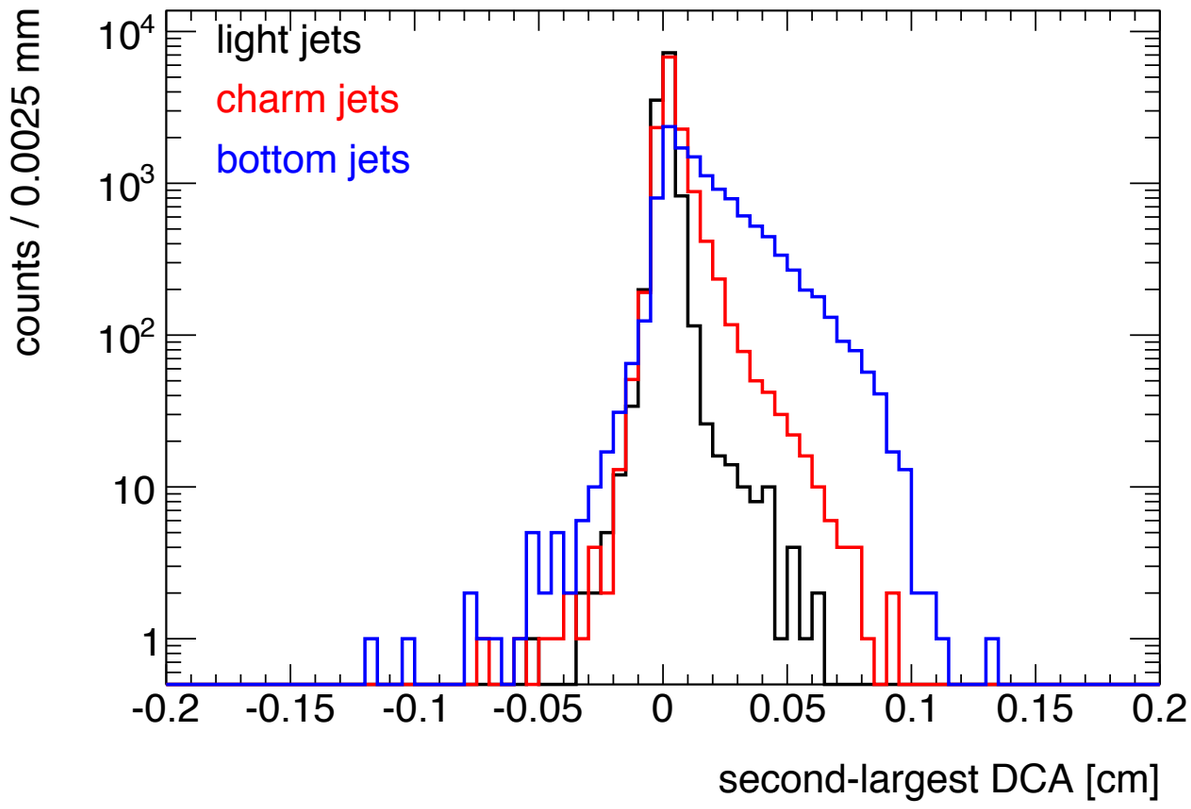




- Left: distribution over number of reco tracks with p_T > 0.5 GeV with ΔR < 0.4 to a truth jet
  - ➡ with MIE-proposal tracking, plenty of tracks to cut on
  - ➡ can see *b*-jets have larger multiplicity
- Right: same distribution but for p_T > 1 GeV

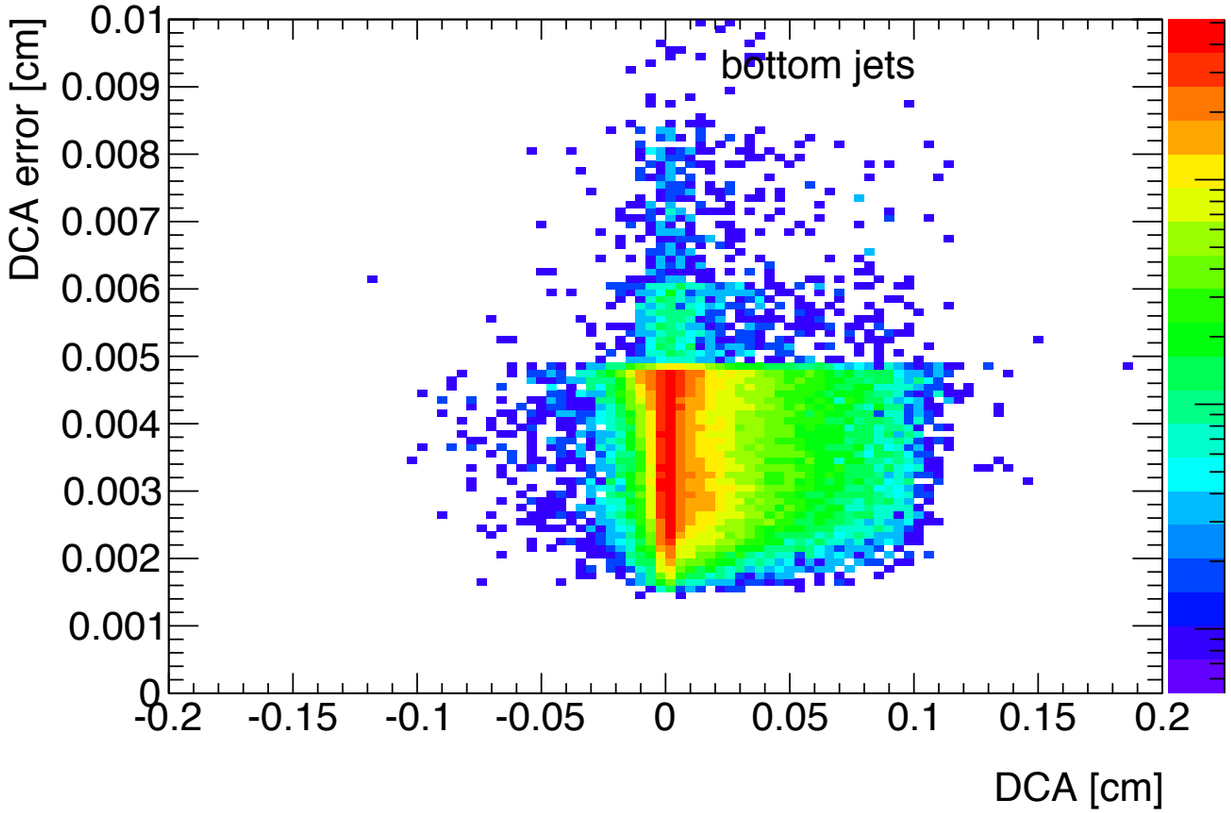
Largest- and 2nd-largest-DCA

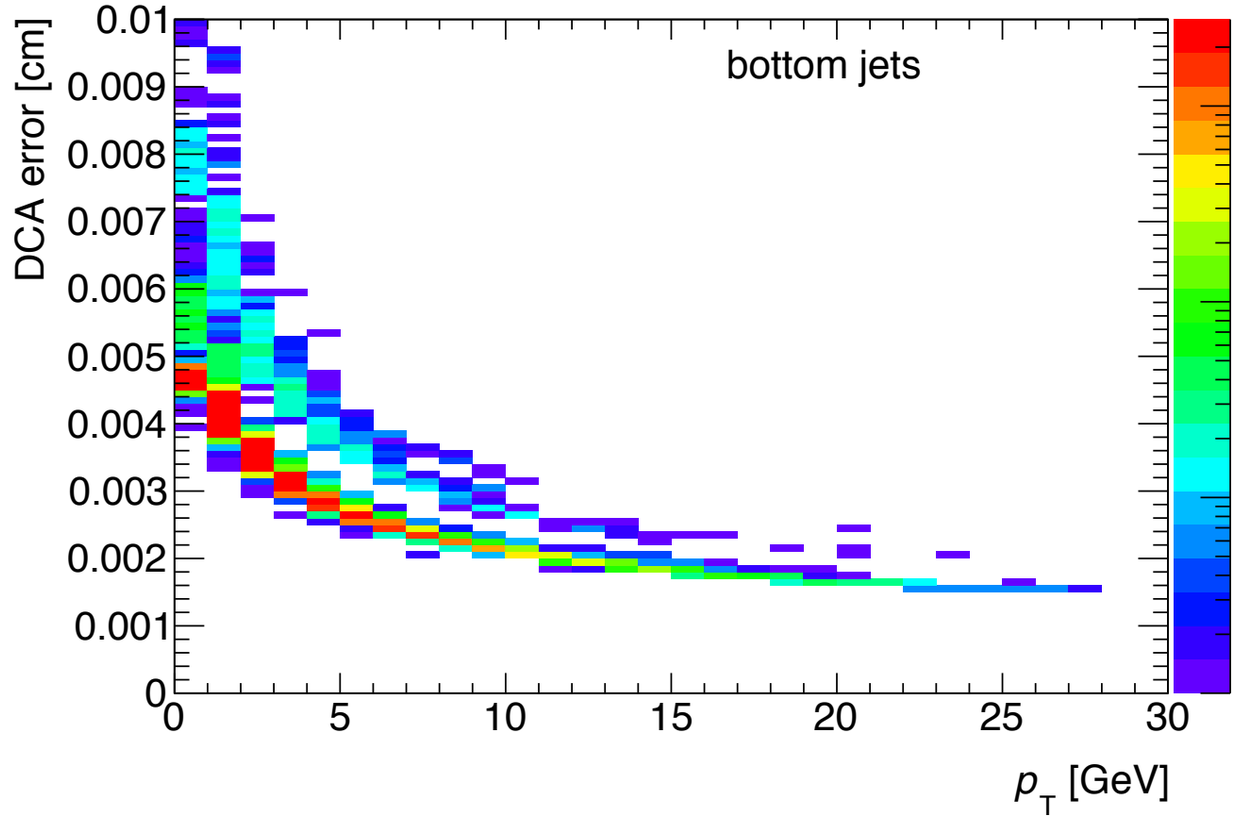




- Left: largest-DCA of the tracks in jet, by jet flavor
  - ➡ as expected, prominent tails in charm/bottom jets, but also non-trivial tail in light jets (partially from small b/c admixture)
- Right: second-largest-DCA among tracks in jet, by jet flavor
  - ➡ still large tail in bottom jets, but suppressed in light jets

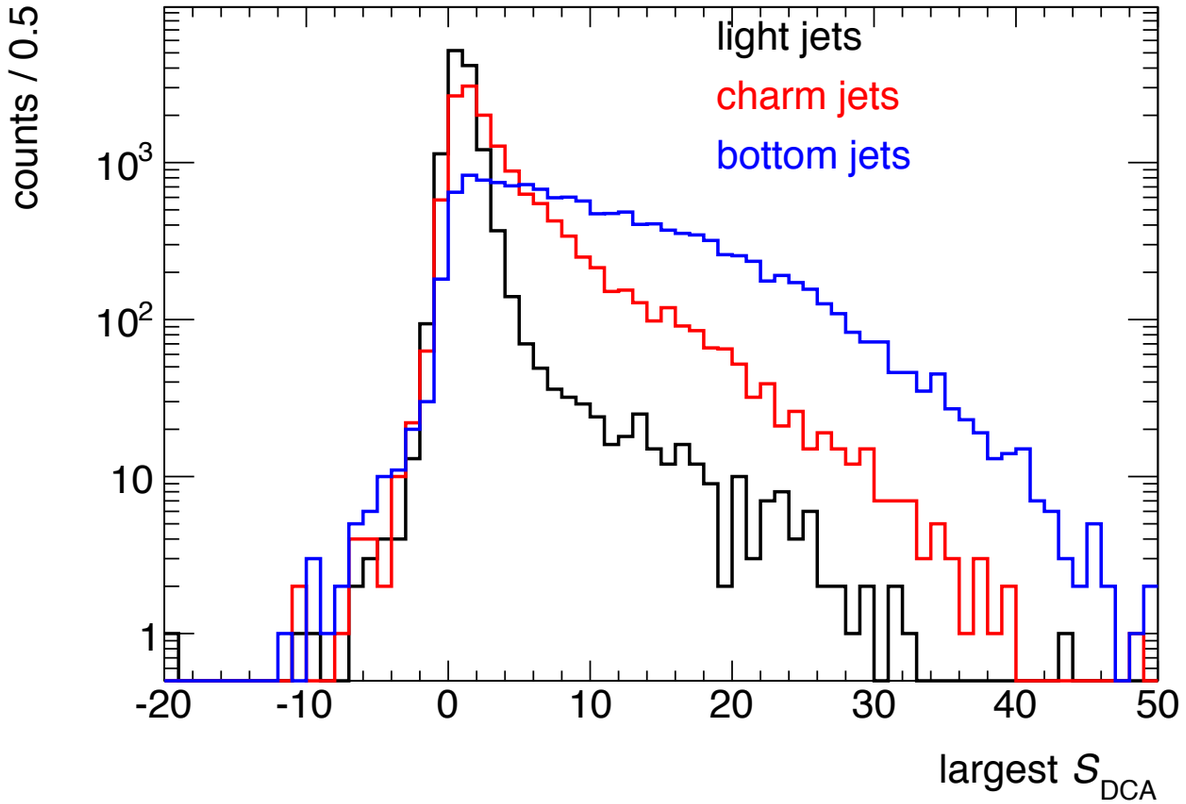
DCA uncertainty

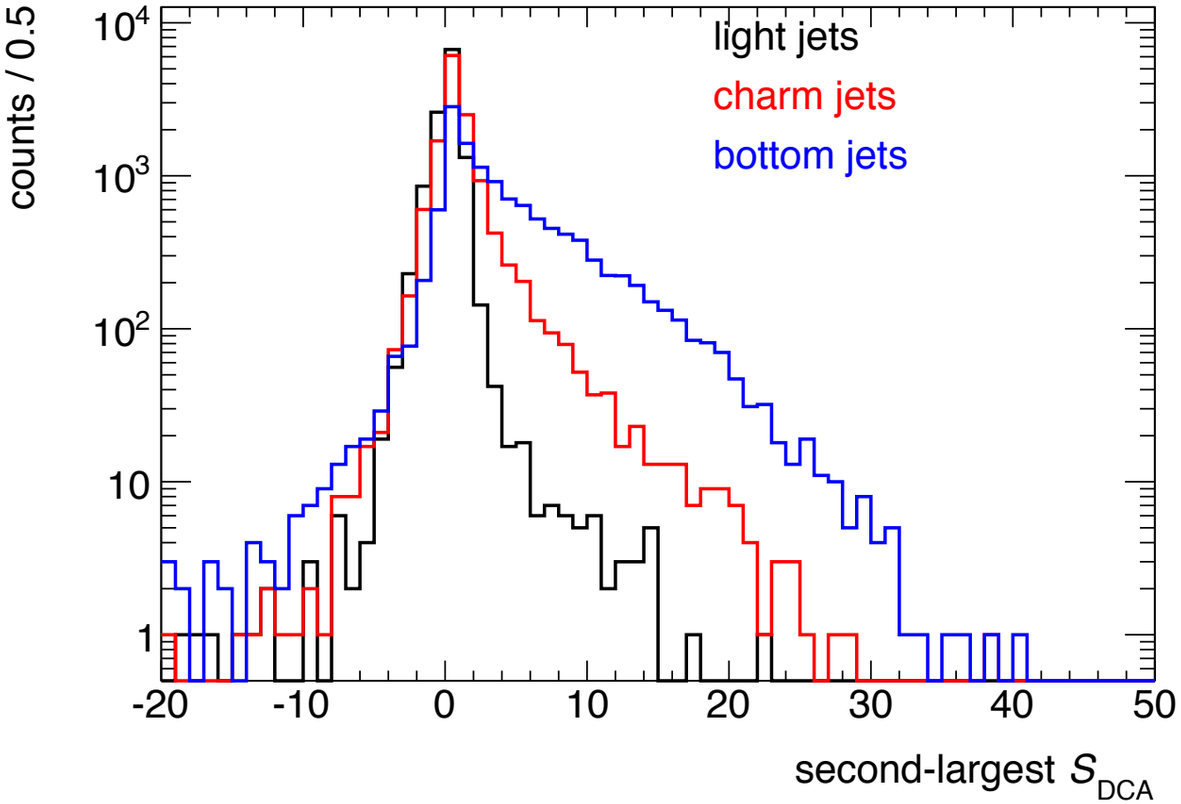




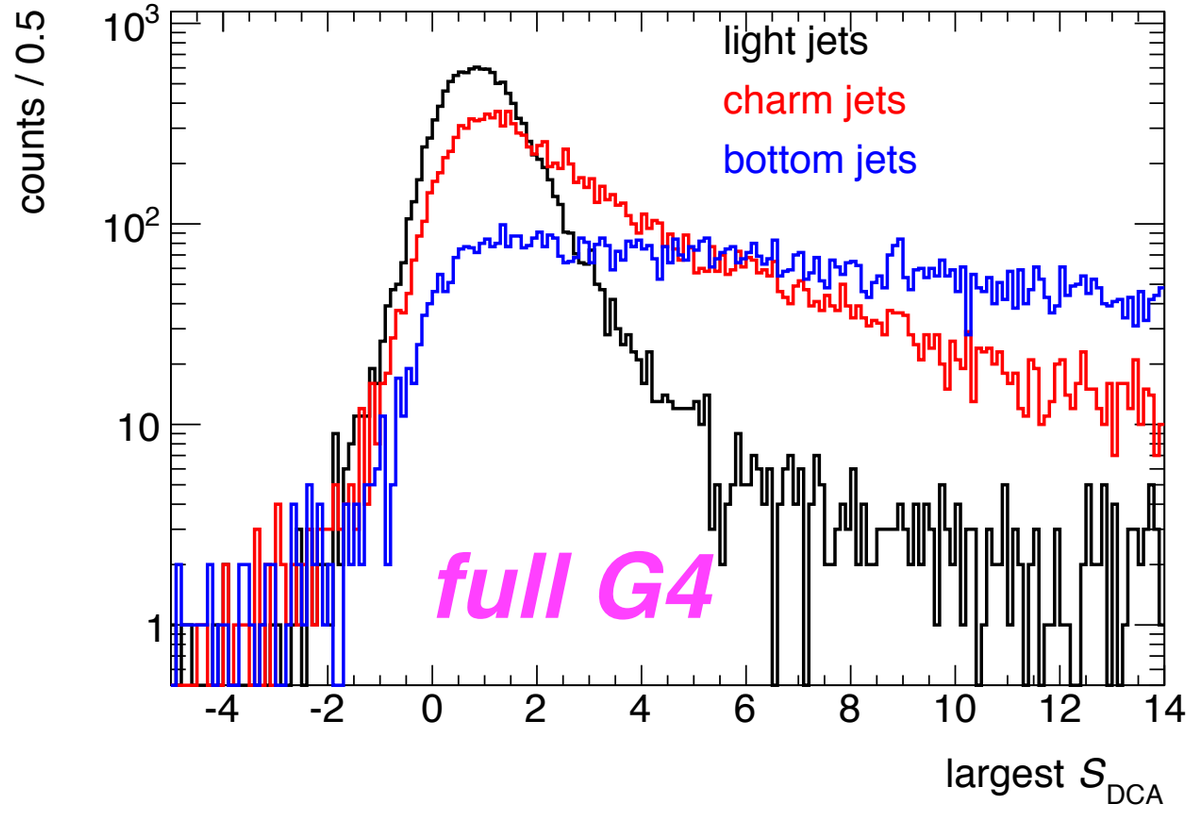
- Uncertainty on DCA (->get_dca2d_error()) for tracks in bottom jets, vs. track DCA (left) and track p_T (right)
- As an alternative, categorize tracks by S_{DCA} = DCA / DCA error
  - ➡ gives larger discriminating power to high-p_T tracks with a more-precisely-known DCA
  - ➡ for the MIE update, we used TrackCounting with S_{DCA}, not DCA

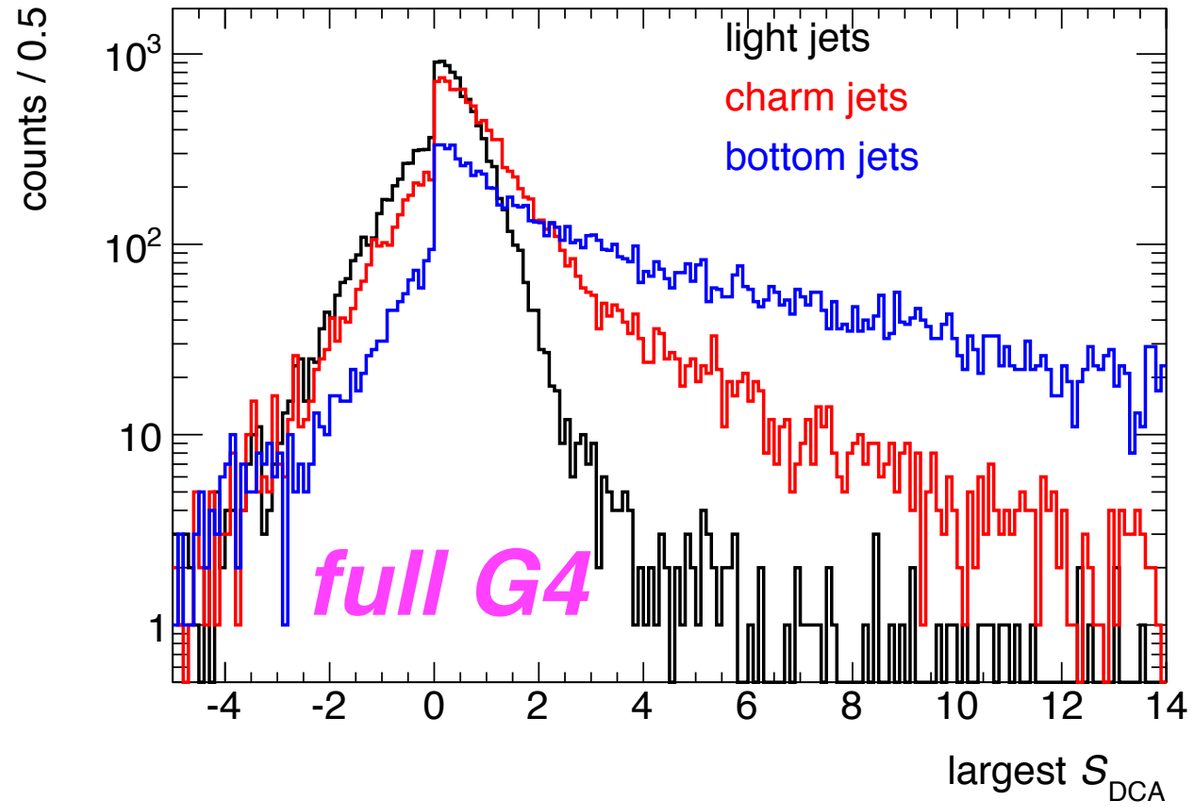
Largest- and 2nd-largest-S_{DCA}

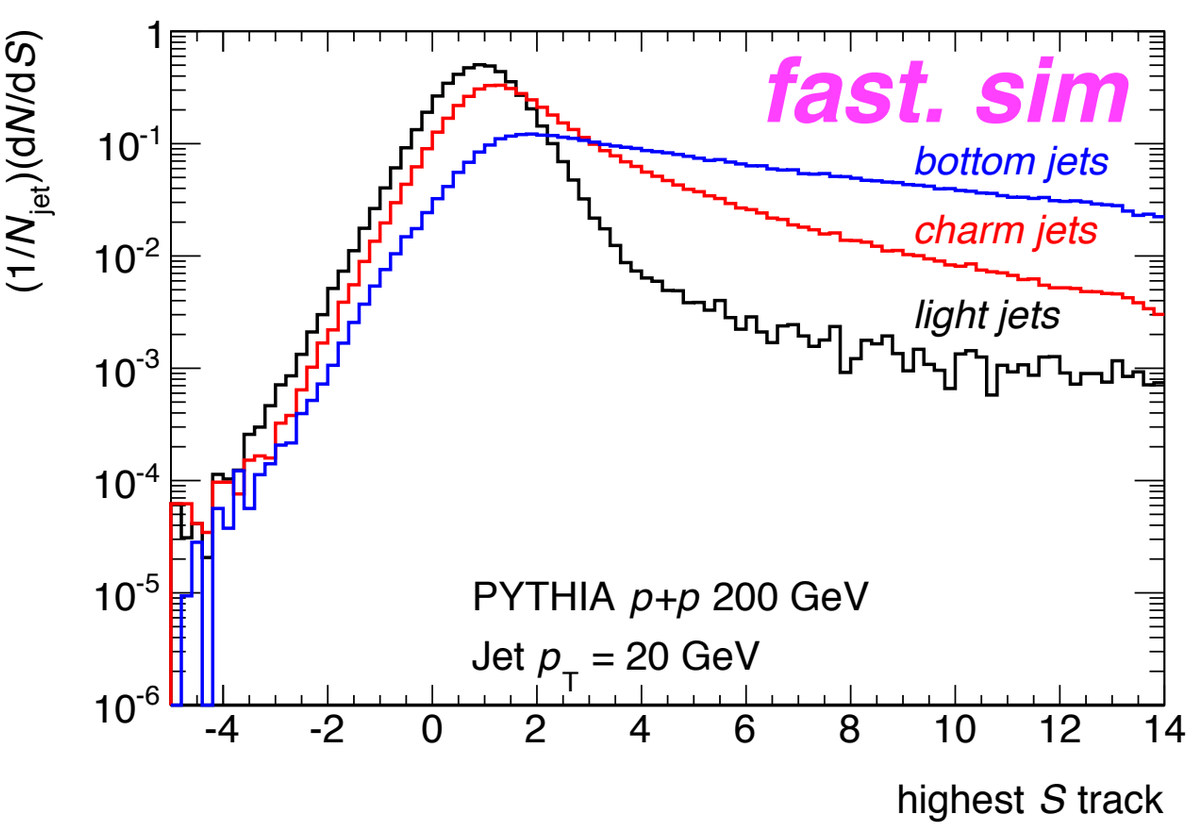


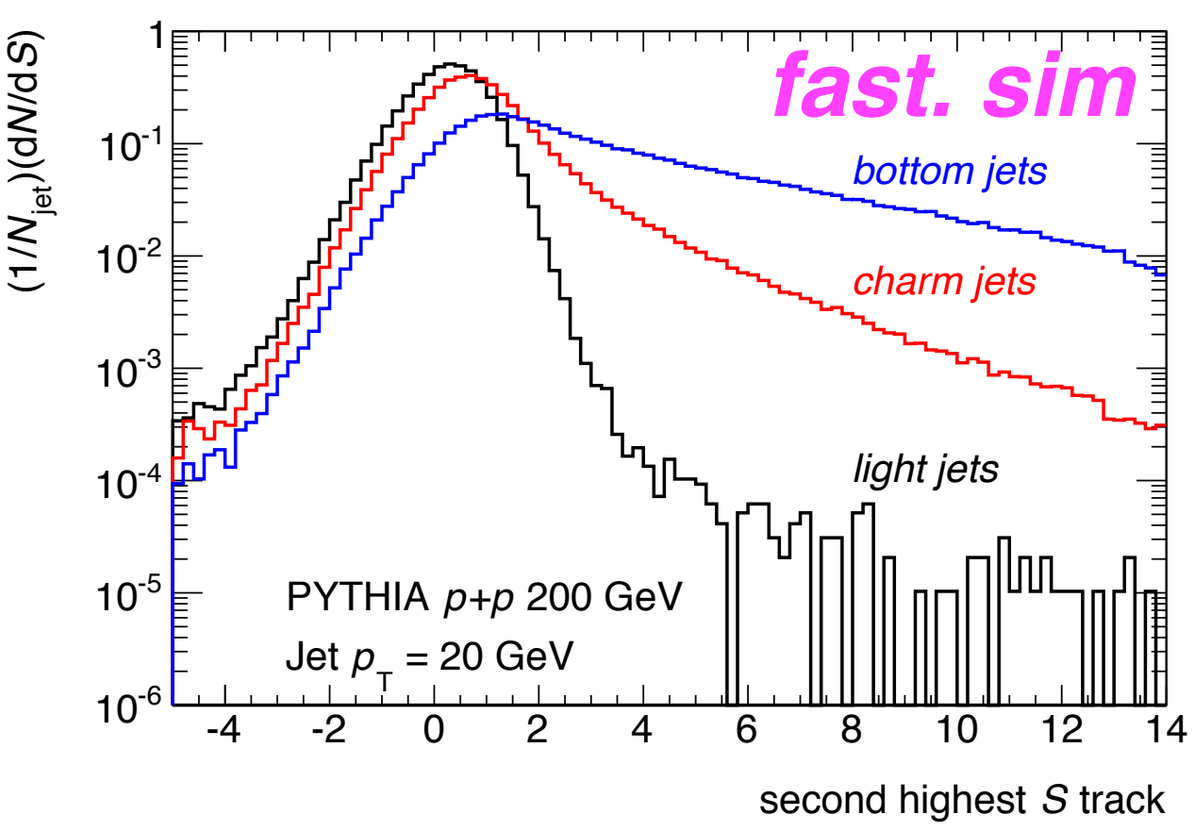


- Left: largest-S_{DCA} of the tracks in jet, by jet flavor
- Right: second-largest-DCA among tracks in jet, by jet flavor



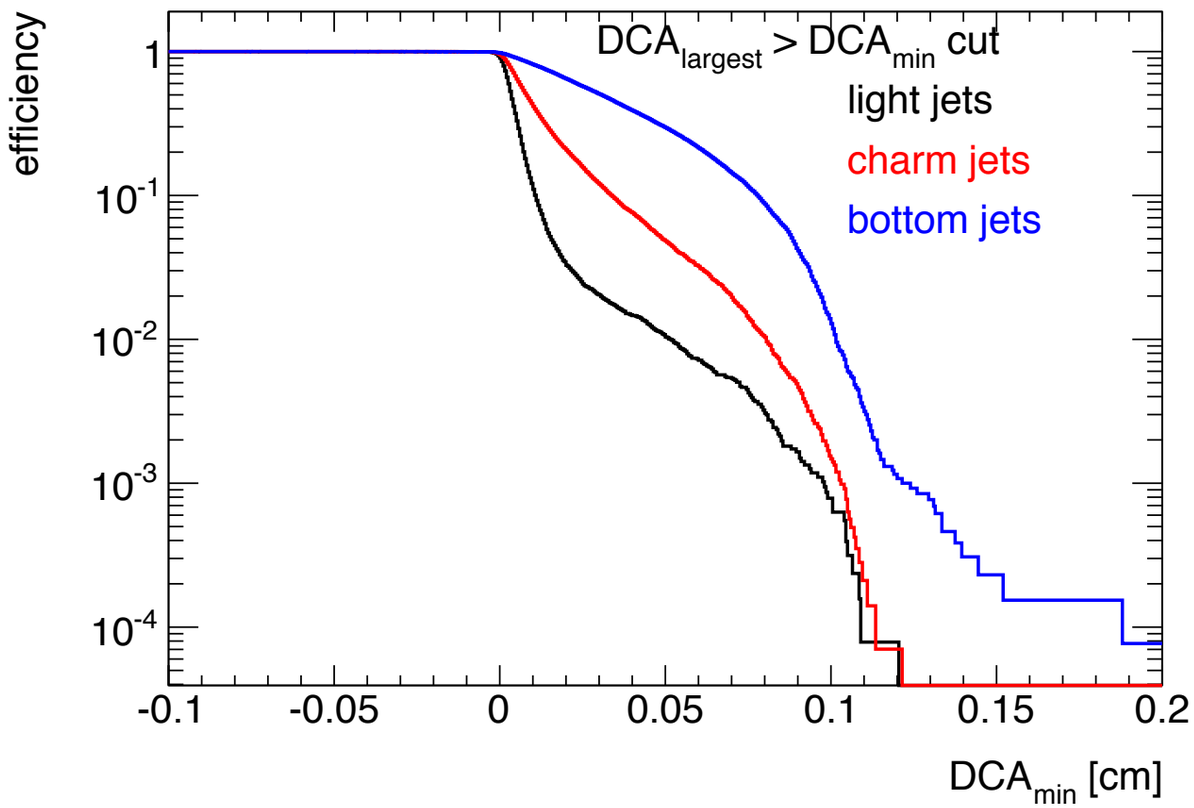


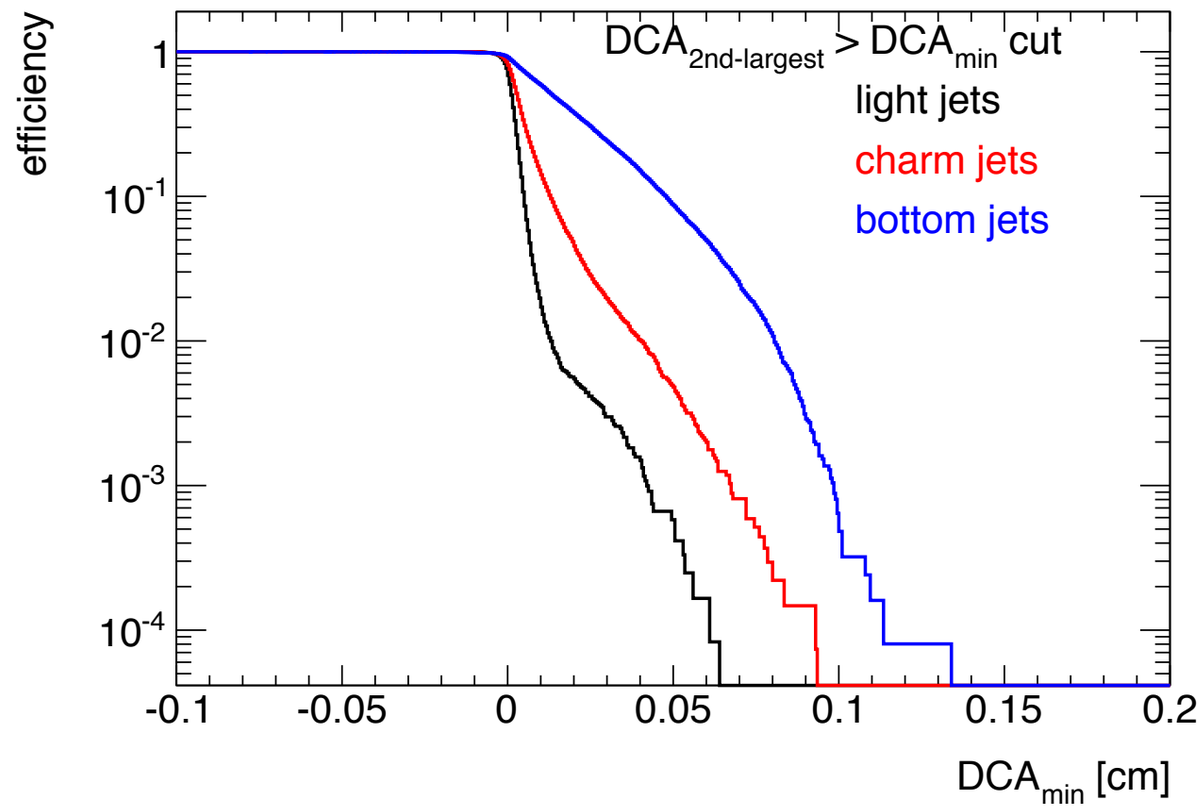




- Zoom in on *x*-range to compare to MIE proposal update plots

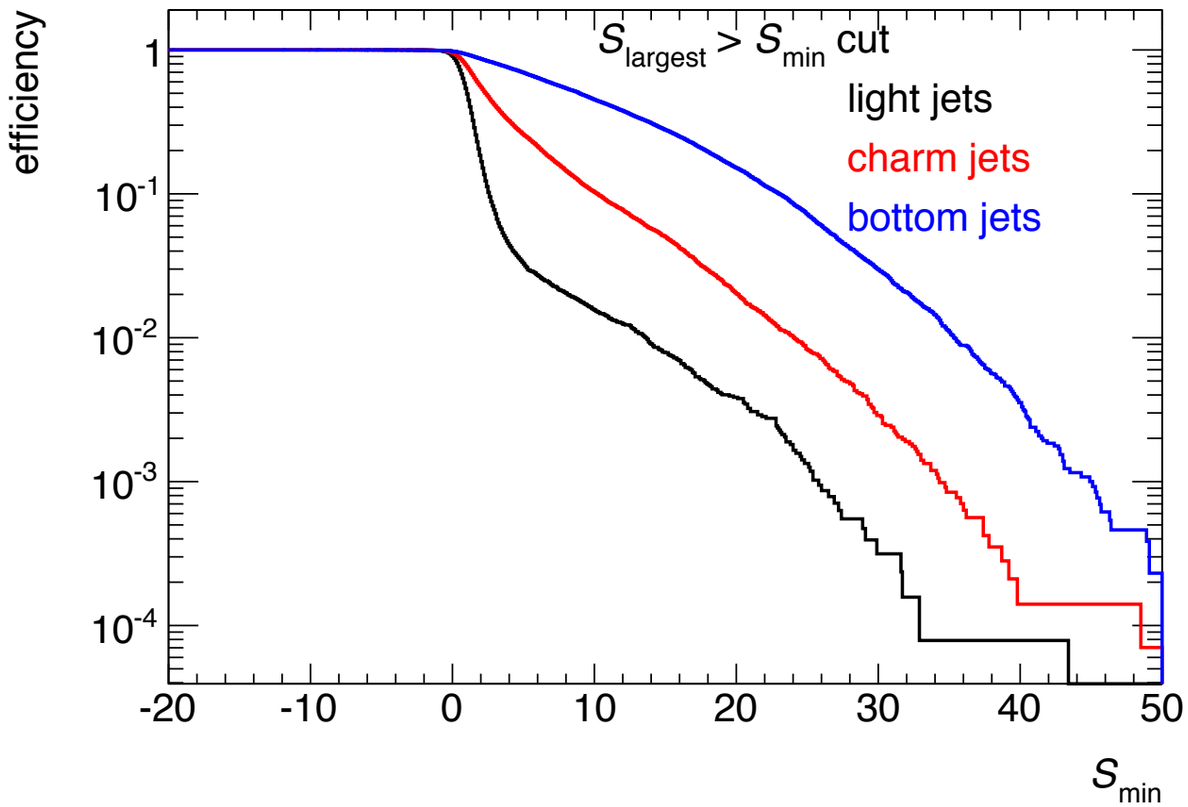
Efficiencies for min. DCA cuts

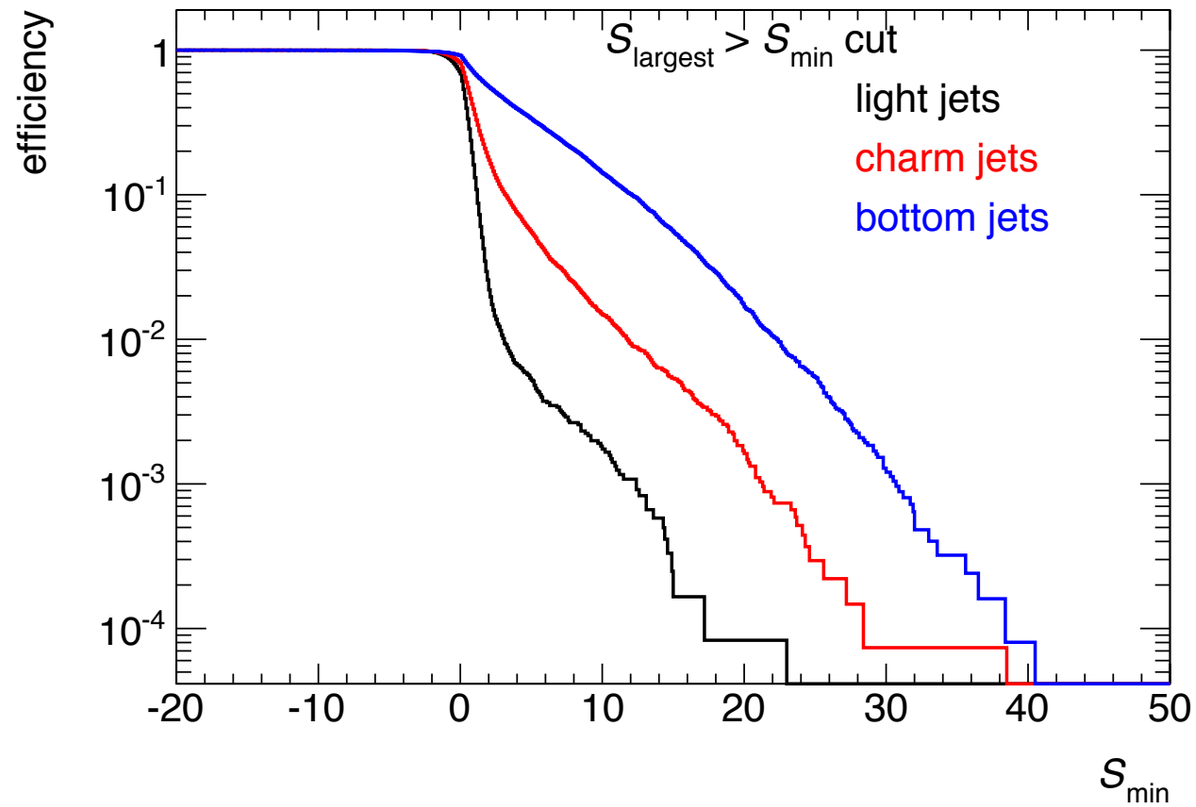




- Now: require jets to have a largest or second-largest DCA above some minimum value DCA_{min}
  - ➡ What is the efficiency $\in$ for this cut for the different flavors, as a function of DCA_{min}?
- Above: $\in$ for requiring largest or second-largest DCA > DCA_{min}
  - ➡ $\in$ decreases with stricter cut, and is ordered by flavor

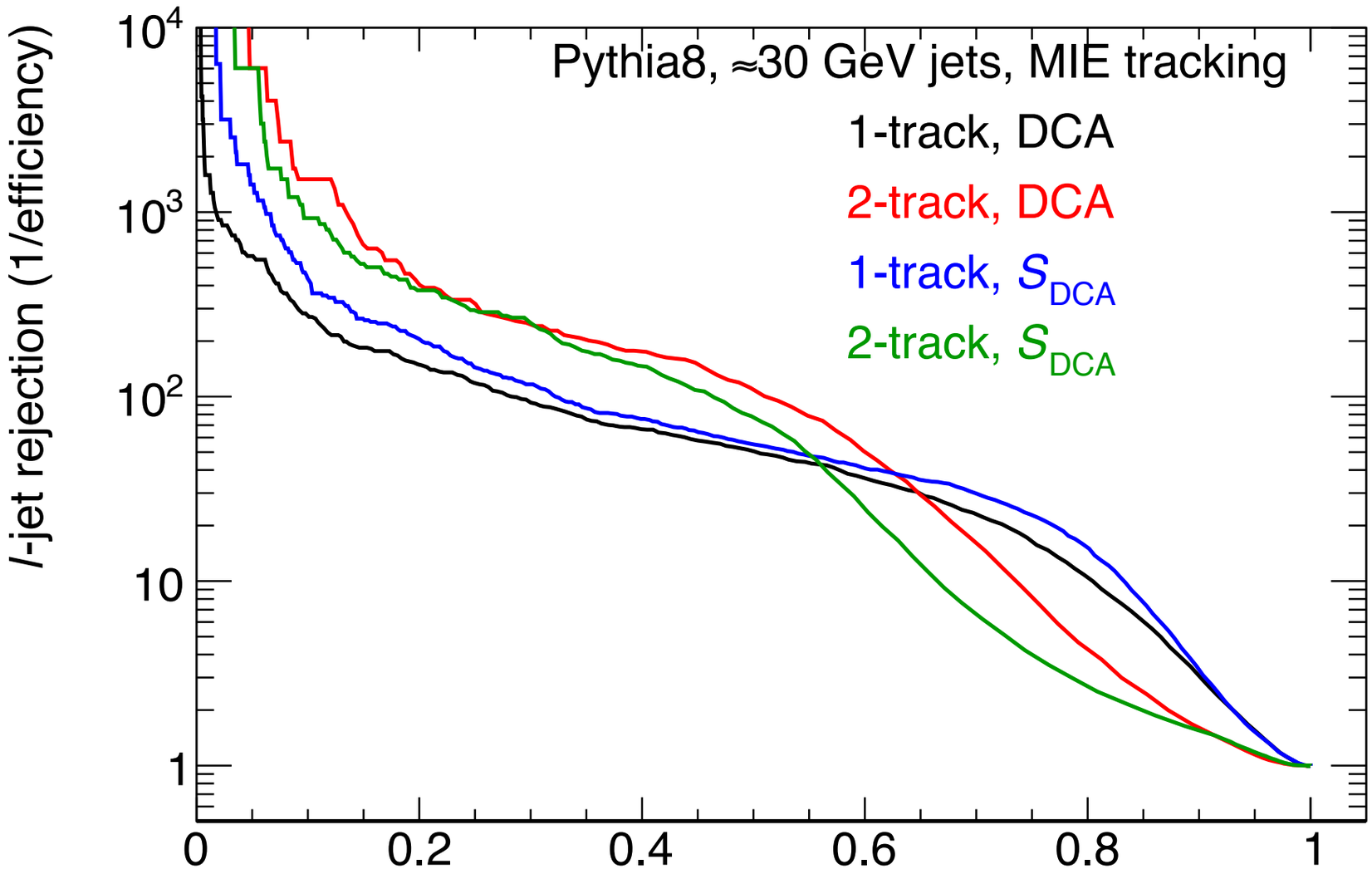
Efficiencies for min. S_{DCA} cuts





- Now: require jets to have a largest or second-largest S_{DCA} above some minimum value S_{min}
- Above: $\in$ for requiring largest or second-largest $S_{DCA} > S_{min}$

b-jet eff vs. *l*-jet rej.



- Important caveats:
  - ➡ my "*b*-jets" are directly produced bbar pairs (no FE, GS production modes)
  - ➡ my "*l*-jets" are actually inclusive jets and so have a small admixture of *b*'s and *c*'s
- Will have to fix these before taking results seriously

Conclusion

- First look at *b*-tagging with full G4 tracking simulation of ideal seven layer silicon in *pp* collisions

  ➡ caveat: need more appropriate *l*/*c*/*b*-jet event samples before extracting a purity/efficiency curve

- Ready to simulate other tracking configurations as part of response to Berndt's charge

  ➡ for example, can investigate effects of inefficiency/ganging in reconfigured VTX pixel + strips…